



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The limits of predictability of volcanic eruptions from accelerating rates of earthquakes

**Citation for published version:**

Bell, AF, Naylor, M & Main, IG 2013, 'The limits of predictability of volcanic eruptions from accelerating rates of earthquakes', *Geophysical Journal International*, vol. 194, no. 3, pp. 1541-1553.  
<https://doi.org/10.1093/gji/ggt191>

**Digital Object Identifier (DOI):**

[10.1093/gji/ggt191](https://doi.org/10.1093/gji/ggt191)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Geophysical Journal International

**Publisher Rights Statement:**

Published in Geophysical Journal International by Oxford University Press. Authors retain copyright (2013)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## The limits of predictability of volcanic eruptions from accelerating rates of earthquakes

Andrew F. Bell, Mark Naylor and Ian G. Main

*School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK. E-mail: a.bell@ed.ac.uk*

Accepted 2013 May 8. Received 2013 May 2; in original form 2013 February 12

### SUMMARY

Volcanic eruptions are commonly preceded by increased rates of earthquakes. Previous studies argue that in some instances these sequences follow the inverse Omori law (IOL) and that this model could be the basis for forecasting the timing of eruption onset. However, the catalogue of pre-eruptive sequences is small, and the performance of the IOL as a forecasting tool remains largely untested. Here, we use simulations to quantify upper limits to the accuracy and bias of forecast eruption times based on the IOL in the ‘best-case’ scenario that uncertainty only arises from model parameter estimation from single realizations of a stochastic point process. We compare different methods for forecasting based on the IOL, and demonstrate that a maximum-likelihood method yields more accurate and less-biased forecasts than methods currently employed. Even in these idealized conditions, we find that large forecast uncertainty and false alarms are inherent features of the mathematics of the IOL. For example model parameter values and 500-d pre-eruptive sequence durations, at 25 d before the eruption, 10 per cent of the forecasts are more than 8 d early or late if the power-law exponent is known *a priori*, and more than 18 d early or late if the power-law exponent is unknown. We also evaluate methods for model comparison and estimation of the power-law exponent. These techniques are applied to examples of real pre-eruptive earthquake data sets. We find evidence for systematic deviations from the idealized model, indicating the action of multiple processes and resulting in greater forecast error than in the synthetic examples, especially close to the eruption time.

**Key words:** Time-series analysis; Volcano seismology; Statistical seismology.

### 1 INTRODUCTION

The development of falsifiable, quantitative methods for forecasting eruptions is an important challenge facing volcanology. Key aspects of volcanic activity that are the target of forecasts include whether an episode of unrest will lead to eruption or not, the onset time of the eruption and the size, style and location of the eruption. Future forecasting strategies will involve a probabilistic framework (Marzocchi & Bebbington 2012), incorporating empirical statistics and expert opinion, along with information from predictive physics-based models that quantitatively relate the temporal evolution of precursory geophysical signals to the probable timing of a future eruption. However, before the use of a forecasting model can be justified, its potential forecasting performance must be quantified, and validity tested against real data sets. In addition to the forecasting benefits, testing the performance of competing models may provide a better understanding of the physical processes acting at volcanoes.

In this paper, we focus on methods to forecast the timing of a forthcoming eruption based on accelerating rates of earthquakes. Many types of volcanic eruptions and associated phenomena are

preceded by increases in the rate of earthquakes, and these signals are a key piece of information used in volcano monitoring (Sparks 2003; McNutt 2005; Chouet & Matoza 2013). Signals involving volcano-tectonic (VT) earthquakes have been interpreted as reflecting changing stress conditions in the volcano, for example, due to the emplacement or movement of magma (Roman & Cashman 2006), or material damage of the volcanic edifice through fracture growth and fault movement (Kilburn 2003; Lengliné *et al.* 2008). Similar rate changes have been identified in low-frequency earthquakes, associated with changes in magma ascent velocities (Hammer & Neuberg 2009) and magma failure due to elevated strain rates in the magma column (Neuberg *et al.* 2006). Forecasts of the onset time of eruptions on the basis of these data have proved reasonably effective in practice, though performance has not been formally assessed. The challenge of forecasting the size and style of the event is, perhaps, a greater one, and will involve additional consideration of factors including signal source mechanism, and magma and rock properties and interactions.

A range of different deterministic models have been proposed to relate temporal changes in the rate of earthquakes to the onset time

of the eruption. Although based on a variety of different paradigms, many of these approaches argue that (under conditions of constant or slowly increasing stress) the rate of earthquakes will increase with time towards the onset of eruption according to a power law (Tokarev 1971; Voight 1988; Kilburn & Voight 1998; Main 1999; De la Cruz-Reyna & Reyes-Davila 2001). Such an increase is described by the inverse Omori law (IOL)

$$\lambda(t) = \frac{k}{(t_e - t)^p} \quad (1)$$

where  $\lambda$  is the rate of earthquakes at time  $t$ ,  $t_e$  is the time of the onset of the eruption,  $k$  is the amplitude (daily rate of earthquakes 1 d before eruption) and  $p$  is the power-law exponent. Formally, eq. (1) is an exact solution for the case of constant stress (creep loading). Solutions are also available for slowly increasing stress, for example, due to a monotonic increase in magma pressure. However, the result can be fitted also with an inverse power law of the form (1), albeit with a different exponent  $p$  (Main 2000). In this simple form, the IOL has been the basis for many retrospective analyses of eruption precursors, and has been widely proposed as a potentially important component of eruption forecasting strategies (Voight 1988; Reyes-Davila & De la Cruz-Reyna 2002; Sparks 2003). The IOL model for earthquakes before volcanic eruptions has been applied to data from a variety of different types of eruption (Voight 1988; Smith *et al.* 2007; Bell *et al.* 2011a). It has also been applied to signals that are likely to be associated with a range of volcanic phenomena, such as shallow magma chamber inflation (Smith & Kilburn 2010), dyke emplacement and magma ascent (Smith *et al.* 2007; Hammer & Neuberg 2009), and lava dome failure and onset of explosive eruptions (De la Cruz-Reyna & Reyes-Davila 2001; Lavallee *et al.* 2013).

It is also important to note that the IOL describes the expected evolution of rates of earthquakes associated with deforming crust or magma. Forecasts based on the IOL involve estimates of  $t_e$ , the time at which the rate of earthquake will tend towards infinity. Forecasts of the timing of an eruption need to be conditioned by the probability that  $t_e$  coincides with an eruption. Similar trends should be expected, and have been observed, before changes in eruption style, magma intrusion events or other non-eruptive phenomenon.

Despite the widespread interest in the IOL, there have only been a few studies that have tested its performance as a forecasting tool on volcanic data, and most of these have only looked at individual case studies (De la Cruz-Reyna & Reyes-Davila 2001; Smith & Kilburn 2010; Bell *et al.* 2011b). Some studies have also tested forecast performance on laboratory data (Lavallee *et al.* 2008; Tuffen *et al.* 2008; Bell *et al.* 2011b), and experimental data sets promise much potential for future improvements of forecasting models and tests. However, a key question that remains unanswered is: how good are forecasts of eruption onset time based on the IOL? The concept of forecast ‘goodness’ is multifaceted; here we focus on goodness in terms of ‘quality’—the correspondence between the forecasts and the matching observations (Murphy 1993). Forecast quality is dependent on properties of the data, properties of the forecasting models and the nature of the relations between the data and the forecast. Reduced forecast quality will result from both epistemic factors (e.g. imperfect forecasting models) and aleatoric factors (e.g. inherent randomness in the data). Pre-eruptive processes are varied and complex and it is likely that multiple processes, potentially acting at different timescales, occur simultaneously. Earthquakes may originate in both rock and magma, and may be driven by tectonic stresses, or magmatic and fluid phase pressures. Driving stresses

may fluctuate during the pre-eruptive phase, resulting in earthquake rates that deviate from the simple trends predicted by physical models. Likewise, multiple pressure sources (such as a shallow inflating chamber, an overpressured dyke and gas pockets) may coexist, both triggering earthquakes in the surrounding crust and magma (Gudmundsson 2006, 2012). It is therefore highly unlikely that the IOL perfectly explains the evolution of earthquake rates before all, or perhaps even any, eruptions. Such deviations will result in discrepancies between the forecast and observed eruption times.

Even in the scenario that the IOL does in fact perfectly explain pre-eruptive rates of earthquakes, discrepancies between the forecast and observed eruption times should still be expected. Forecasts based on the IOL involve estimating the model parameters from a partial pre-eruptive earthquake sequence. Given the stochastic nature of the earthquake process, a range of different parameter values (including  $t_e$ ) may adequately explain the data. Although the true parameter values should be within this range, there will be inherent uncertainty in the forecast eruption time. The range of viable parameter values will decrease as more data become available, reducing the uncertainty in the forecast. Observational errors (in the magnitude, timing, location and source of earthquakes) will increase the uncertainty in the forecast. Comprehensive quantification of these uncertainties and their impact on the quality of subsequent eruption forecasts will require truly prospective (before the eruption has happened) testing on many different pre-eruptive data sets over many years. This quantification should be a long-term goal of volcanology. In the meantime, it is possible to address parts of this problem using either retrospective analyses of archive data or simulated prospective analyses of synthetic data.

Here, we first use simulations to quantify an upper limit to the potential forecasting performance of the standard IOL as commonly applied to volcanic earthquake data. In this ‘best-case’ scenario, we restrict the uncertainty to that associated with the estimation of model parameters from a pre-eruptive earthquake catalogue. Large numbers of synthetic earthquake catalogues are generated where the earthquake rate evolves with time according to the IOL, and where every earthquake sequence culminates in an eruption (i.e. there is no observational or model error). We compare the performances of different methods for forecasting the eruption time by estimating the model parameters from the data. We focus on the bias (the correspondence between the mean forecast and the mean observation) and accuracy (the correspondence between individual pairs of forecasts and observations) of forecasts (Murphy 1993), quantifying the temporal evolution of these metrics. We then examine the performance of these techniques on real pre-eruptive earthquake data sets, using the performance against synthetic examples as benchmarks to identify behaviour that is inconsistent with the IOL and indicative of the presence of additional or superposed processes.

## 2 METHODS FOR FORECASTING ERUPTIONS USING THE IOL

When making a forecast based on the IOL, the eruption time,  $t_e$  in eq. (1), and other model parameters are estimated from the times of earthquakes occurring in the pre-eruptive sequence. As each pre-eruptive sequence of earthquakes is a unique realization of an underlying stochastic process, we should expect there to be an inherent uncertainty associated with these parameter estimates even if the IOL perfectly explains the data. Different statistical methods have been proposed to estimate the parameters of the IOL from earthquake occurrence times.

The Failure Forecast Method (FFM; Cornelius & Voight 1994) relies on linearizing the IOL in the form

$$(\lambda(t))^{-\frac{1}{p}} = k^{-\frac{1}{p}} (t_e - t). \quad (2)$$

The value of  $p$  is assumed to be known *a priori*, or is estimated from the data through other means. The eruption time is then determined using standard least-squares regression on average earthquake rates in discrete bins, with the number of bins chosen somewhat arbitrarily. Commonly,  $p \approx 1$ , in which case the solution involves a regression of inverse rate against time.

Bell *et al.* (2011b) demonstrate that the FFM provides a biased, imprecise estimate of  $t_e$  for both pre-eruptive earthquake and strain data. The FFM fails to account correctly for the error structure of the data: specifically the linearization operation invalidates the Gaussian assumption underlying least-squares regression. Instead, Bell *et al.* (2011b) propose a regression technique based on a generalized linear model (GLM; Nelder & Wedderburn 1972). The GLM technique is a generalization of linear regression that allows for error distributions that are non-Gaussian, and a ‘link function’ to provide the relation between the linear model and dependent variable. Using the Poisson error distribution associated with earthquake rate data, and a power-law ‘link function’ to relate the independent and dependent variables, the GLM can be used to apply the IOL (Bell *et al.* 2011b). However, the GLM still requires regression to be carried out on incremental earthquake frequencies (‘binned’ data) rather than analysing the full point process, leading to a possible residual bias in forecasting the eruption time.

An alternative parameter estimation method is to use maximum-likelihood (ML) estimation. The ML parameters are those that result in a model that gives the observed data the greatest probability, and are those that maximize the likelihood function. For data that have a Gaussian distribution, least-squares regression provides the ML estimate. However, this is not the case for point processes, such as earthquake occurrence. Instead, for many earthquake occurrence models the likelihood function takes a relatively simple form (Daley & Vere-Jones 2005) and can be calculated directly using the occurrence times of earthquakes, eliminating the need to bin the data prior to analysis. The ML model parameters maximize this likelihood function, and are usually established by numerical optimization (often, for computational reasons, performed on the logarithm of the likelihood). Ogata (1983) described the likelihood function for the modified Omori law for aftershock occurrence. The likelihood function for the IOL (eq. 1) takes a similar form. The log-likelihood for the IOL for a catalogue of earthquakes occurring at times  $t_i$  within a given time interval  $(t_0, t_1)$  is given by

$$\begin{aligned} & \log L(k, t_e, p | \{t_1, \dots, t_n\}) \\ &= \sum_{i=1}^n \log(k(t_e - t_i)^{-p}) - \frac{k}{1-p} ((t_e - t_1)^{1-p} - (t_e - t_0)^{1-p}) \end{aligned} \quad (3)$$

for  $p \neq 1$ , and

$$\begin{aligned} & \log L(k, t_e, p = 1 | \{t_1, \dots, t_n\}) \\ &= \sum_{i=1}^n \log(k(t_e - t_i)^{-1}) + k(\log(t_e - t_1) - \log(t_e - t_0)) \end{aligned} \quad (4)$$

for the specific case that  $p = 1$ .

### 3 COMPARING FORECASTING METHODS USING SYNTHETIC DATA

For real data, it is only ever possible to observe one realization of the stochastic process. Using simulations, it is possible to generate a large number of realizations using the same rate function. By taking this approach, it is possible to explicitly quantify the variability in forecast times that simply result from estimating model parameter values from a stochastic point process.

We generate synthetic pre-eruptive earthquake catalogues using a stochastic ‘thinning’ method (Daley & Vere-Jones 2005). Each sequence consists of a unique sequence of event times, and is a single realization of a Poisson process with the rate varying with time according to eq. (1). We do not include earthquake triggering or background seismicity in our simulations. Although these effects are undoubtedly important, they are not accounted for in existing formulations of the IOL forecasting model. Incorporation of these additional parameters in the data and models will only increase the uncertainty in the best-case forecasts. To illustrate our methodology and determine representative forecast uncertainties, we choose illustrative generating parameter values of  $k = 50$ ,  $p = 0.9$  and  $t_e = 500$ . These values are consistent with those we observe for real data (Reasenber & Jones 1989; Schmid & Grasso 2012) and are consistent with values predicted by the theories cited in Section 1. Our examples in Section 4 have ranges of  $0.6 \leq p \leq 1.0$  and  $8.5 \leq t_e \leq 430$ . Volcanic unrest episodes cover a range of durations, from hours to years (McNutt 2002). However, these results are scalable to longer and shorter sequence durations. Fig. 1 illustrates a single realization of the IOL using these parameter values, and the ML IOL model estimated using the earthquakes that have occurred by 450 d through the 500-d sequence. The model is illustrated in the cumulative form to compare against the total number of earthquakes

$$\Lambda(t) = \frac{k}{p-1} ((t_e - t_1)^{1-p} - (t_e - t_0)^{1-p}). \quad (5)$$

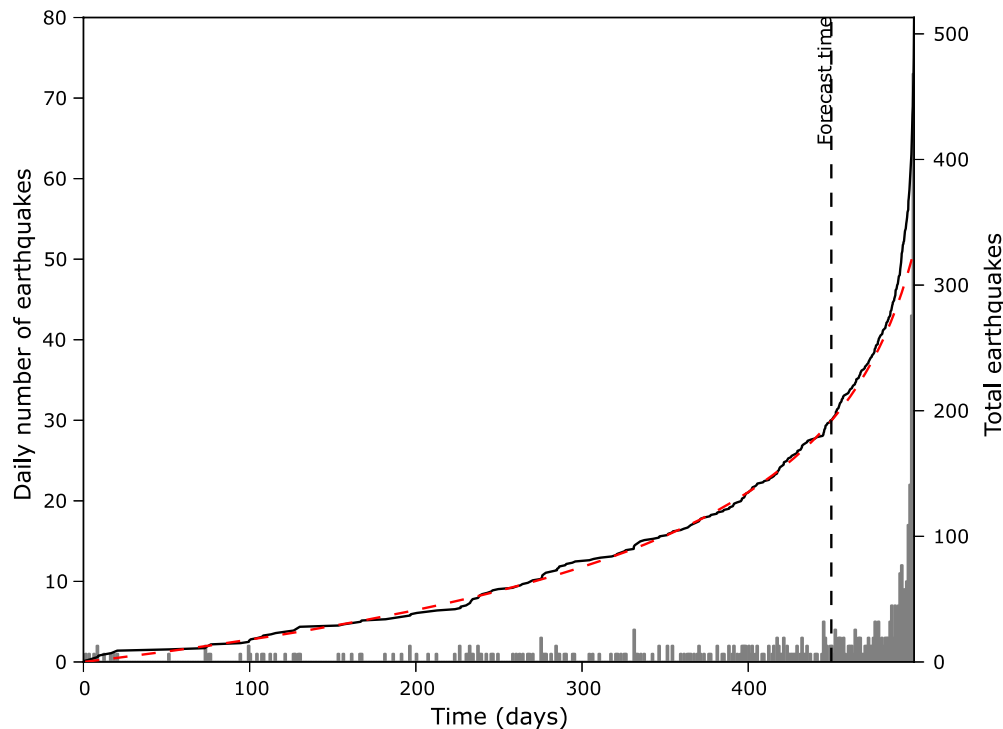
The ML model closely follows the synthetic data used to determine the best fit before 450 d, but systematically deviates from the data when the fit is extrapolated towards the future eruption time.

#### 3.1 Forecasting with known $p$ -value

We begin by assuming that we know the true value of  $p$  *a priori*, and only estimate the productivity and eruption time from the data. This constraint reduces the degrees of freedom in the model and is expected to lead to reduced variability in the forecast eruption times. Theoretical considerations suggest that  $p$  may be expected to take specific values (Main 2000; Kilburn 2003).

We compare the performance of three methodologies for forecasting the eruption time using the earthquake time-series: (1) the FFM (eq. 2); (2) GLM and (3) ML method (eqs 3 and 4). Methods (1) and (2) use earthquake rates determined in 10 equally spaced bins. Method (3) uses the times of individual earthquakes and we find the ML solution by minimizing the negative log-likelihood function using a downhill simplex algorithm. We choose random values for the optimization starting parameters (within a reasonable range), but the optimization is robust to variation in these values.

For each of 500 synthetic catalogues, we apply the three different forecasting methods to estimate  $t_e$  at each time  $t_1$  on the  $x$ -axis (Figs 2a–c, left-hand panels). Forecasts are based only on the times of earthquakes that have occurred in the interval  $t_0, t_1$ , where  $t_0 = 0$ . Each colour thread corresponds to the eruption



**Figure 1.** A single realization of the inverse Omori law with parameter values of  $k = 50$ ,  $p = 0.9$  and  $t_e = 500$ , shown at 90 per cent of the eruption time ( $t_1 = 450$ ). The daily numbers of earthquakes are shown by vertical bars; the total number of earthquakes is shown by the solid black line. Dashed red line illustrates an example model fit, formulated for the total number of earthquakes, and extrapolated to the forecast eruption time.

forecast history associated with a single synthetic catalogue. The range of these estimated eruption times at any given forecast time defines the accuracy of the forecast—a narrow range means a more accurate forecast. For all methods, as time (and the number of earthquake times on which the forecast is based) increases, the accuracy of the forecast increases. The bias in the forecast is defined by the difference between the mean estimated eruption time over all forecasts and the true eruption time. For all models, the mean forecast eruption time tends towards the true eruption time as time increases, indicating a reduction in bias (Fig. 2, right-hand panels). The accuracy of the forecasts is indicated by the 5 and 95 percentiles.

Fig. 3 shows the frequency distribution of forecast eruption times at 85, 95 and 99 per cent of the true eruption time for 2000 synthetic earthquake catalogues using the three different forecasting methods. All methods produce forecasts with a mean that is close to the true eruption time. However: (1) the FFM produces forecasts with a much lower accuracy than the GLM and ML methods; and (2) the GLM method slightly underestimates the eruption onset time as  $t_1 \rightarrow t_e$ . The lower accuracy of the FFM method is because this method fails to account for the true Poissonian error structure of the data. The bias in the GLM method is a result of averaging the earthquake rates in discrete, finite duration, bins. The same effect should be observed in the FFM, but is lost due to the lower accuracy. These properties are consistent with the assumptions underlying the three different methodologies, and are observed for different model parameters and binning intervals.

The statistics of the forecast time distributions provided by the three methods are shown in Table 1. The GLM and ML methods clearly outperform the FFM, which gives a far greater variability in the forecast eruption time (in terms of the average deviation from the mean or true values). For the binned methods (FFM and GLM), the mean forecast tends to a slight underestimate as the eruption time

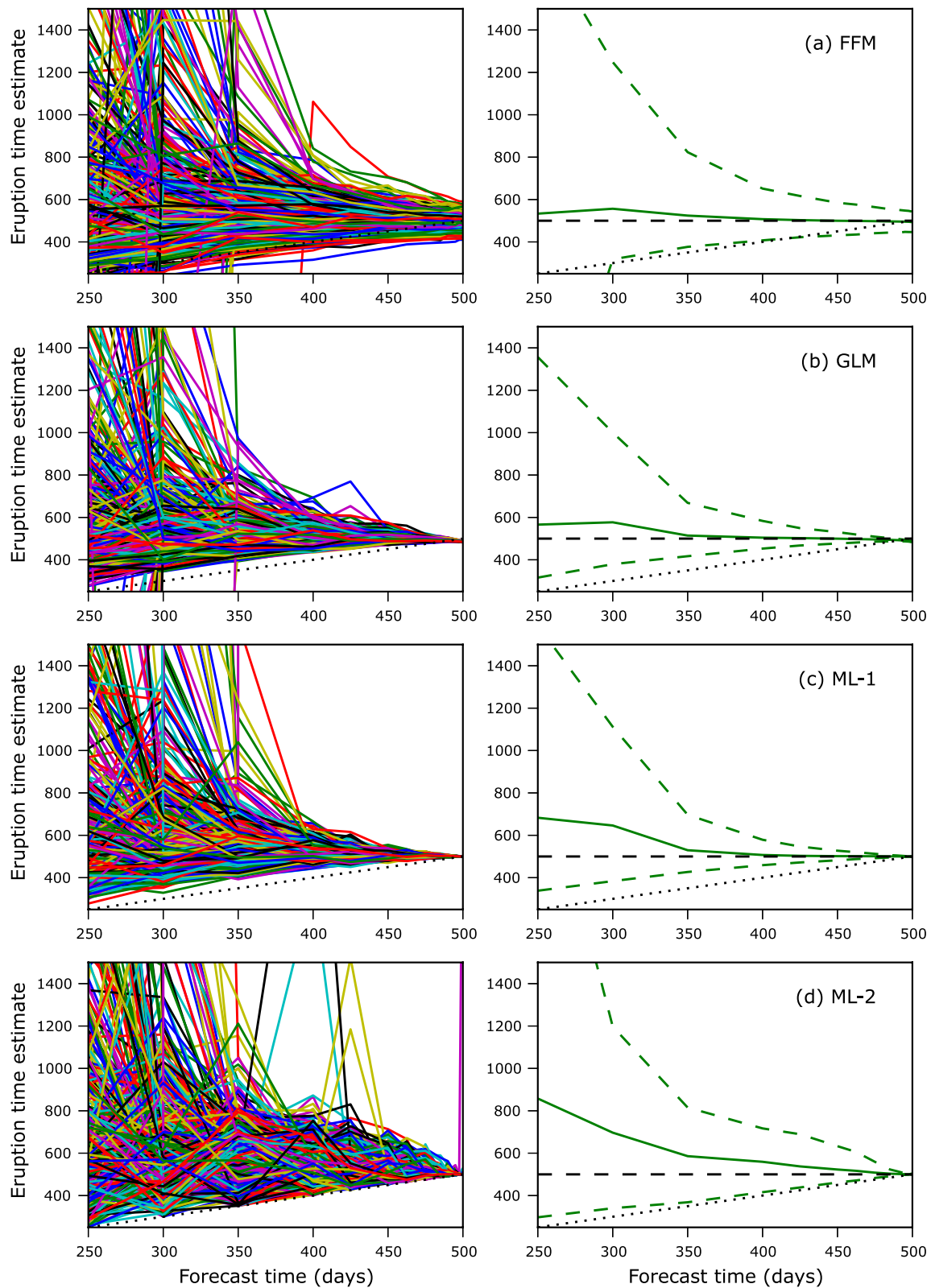
approaches due to the finite bin approximation. For the ML method, at 425 d through a 500-d sequence, 5 per cent of the forecasts will be more than 29 d early and 5 per cent of the forecasts will be greater than 55 d late; at 475 d, 5 per cent of the forecasts will be more than 8 d early and 5 per cent of the forecasts will be greater than 12 d late; at 495 d, 5 per cent of the forecasts will be more than 1 d early and 5 per cent of the forecasts will be greater than 2 d late.

### 3.2 Forecasting with unknown $p$ -value

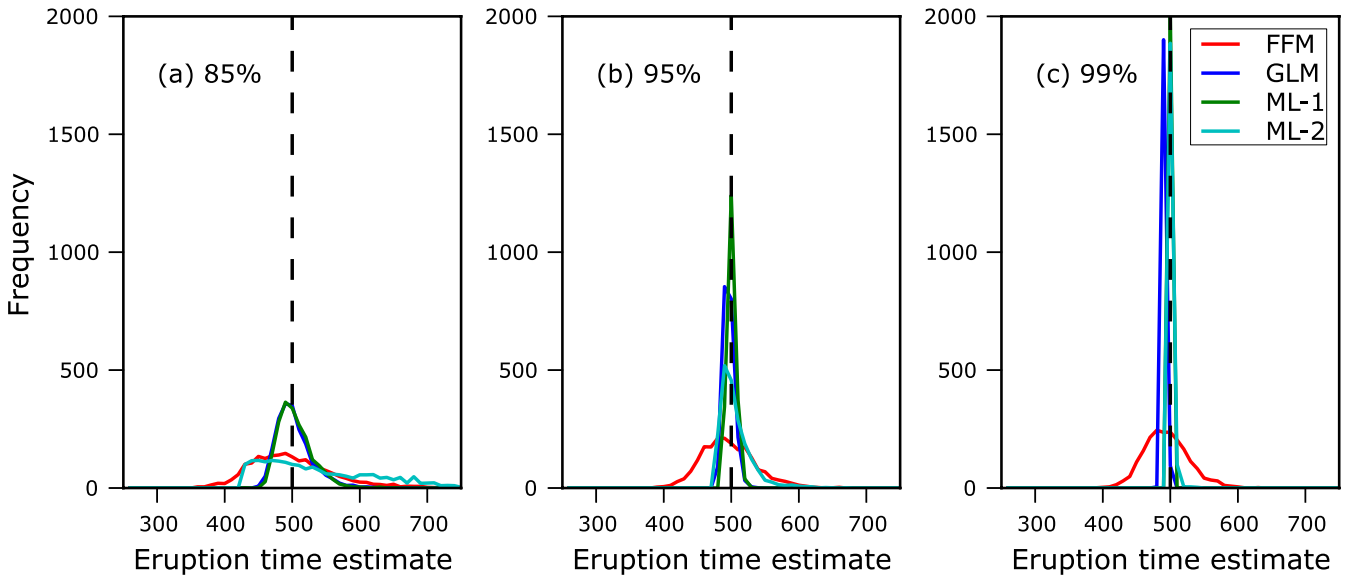
Generally, it is unlikely that  $p$  will be well-known *a priori*, and so the IOL model will have an extra degree of freedom. Although there is no established practice for extending the FFM or GLM methods to a scenario where  $p$  must also be estimated, the ML method can be simply extended. When  $p$  is unknown, the ML method gives stable, repeatable estimates of the IOL model parameters for a reasonable range of random starting parameters.

Fig. 2(d) shows the distribution of forecast eruption times using the ML method where the value of  $p$  is not known *a priori*. Again the forecast mean converges on the true eruption time, and the accuracy increases as the eruption time is approached. However, the accuracy in the forecast is considerably reduced (two to four times) compared to when  $p$  is known. The frequency distributions of forecast eruption times for this ML method are shown in Fig. 3, and the forecast statistics included in Table 1. The corresponding estimates of the values of  $p$  are shown in Fig. 4. Only within the last 50 d of the sequence do the estimates of  $p$  converge on the true solution, even though the eruption time converges much earlier. With shorter duration sequences, this convergence time would be correspondingly later.





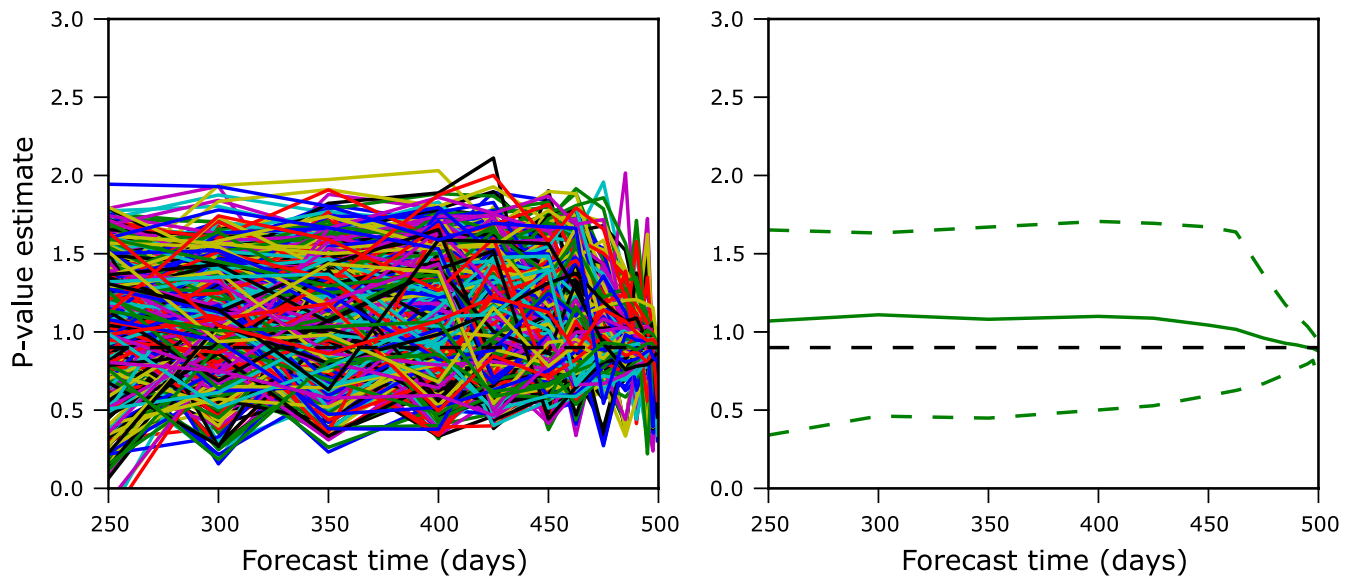
**Figure 2.** Forecast eruption times (left-hand panels) and mean (solid green) and 5 and 95 percentile scores (dashed green; right-hand panels) as a function of time for 500 synthetic earthquake sequences. IOL parameters are estimated using: (a) Forecast failure method (FFM); (b) Generalized Linear model (GLM); (c) Maximum-likelihood with  $p$ -value known (ML1) and (d) Maximum-likelihood method with  $p$ -value unknown (ML2). The true eruption time is shown by the horizontal dashed black line; forecast eruption time =  $t_1$  is shown by the black dotted line for reference.



**Figure 3.** Frequency distribution of forecast eruption times at 85, 95 and 99 per cent (425, 475 and 495 d) of the true eruption time for a 500-d sequence. The four methods to apply the IOL are as for Fig. 2;  $t_e$  is indicated by vertical dashed line.

**Table 1.** Statistics of forecast eruption time frequency distribution at 85, 95 and 99 per cent for  $t_e = 500$  for four different methods. Ogata's ML (1) for the case that  $p$  is known, Ogata's ML (2) for the case that  $p$  is unknown. 5 and 95 per cent correspond to the 5 and 95 percentiles, respectively.

	425 d			475 d			495 d		
	Mean	5 per cent	95 per cent	Mean	5 per cent	95 per cent	Mean	5 per cent	95 per cent
FFM	500	417	607	497	439	563	494	445	551
GLM	503	468	555	496	485	509	491	487	495
Ogata ML (1)	505	471	555	501	492	512	500	499	502
Ogata ML (2)	536	435	684	507	482	554	501	497	505



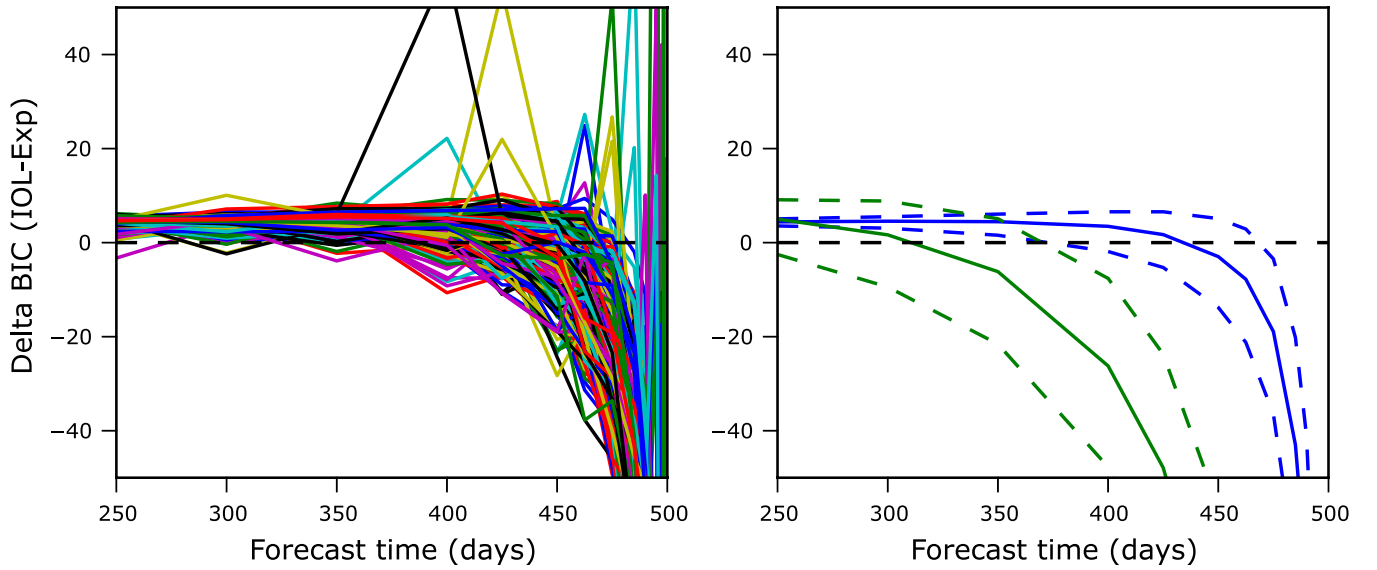
**Figure 4.** ML  $p$ -values (left-hand panel) and mean (solid green) and 5 and 95 percentile scores (dashed green; right-hand panel) as a function of time for 500 synthetic earthquake sequences. The true  $p$ -value is shown by the horizontal dashed black line.

### 3.3 Model comparison

The IOL is the most commonly employed model for the rate of pre-eruptive earthquakes. However, theoretical and observational studies have proposed other models with different degrees of model complexity. In particular, several studies argue that an exponential

model may better explain trends in earthquake rate before some eruptions at basaltic volcanoes (Lengliné *et al.* 2008; Bell *et al.* 2011a; Bell & Kilburn 2012; Kilburn 2012). In the exponential model, the rate of earthquakes evolves with time according to

$$\lambda(t) = k \exp(\phi t), \quad (6)$$



**Figure 5.** Left-hand panel— $\Delta BIC$  (IOL-exponential) as a function of time for 500 synthetic earthquake sequences. Right-hand panel—mean (solid blue) and 5 and 95 percentile scores (dashed blue) of  $\Delta BIC$ .  $\Delta BIC$  (IOL-constant rate) mean and percentiles also shown in right-hand panel for comparison. In both cases, negative values indicate a preference for the IOL.  $\Delta BIC = 0$  is indicated by the horizontal dashed black line.

where  $\phi$  is the exponential constant. Notably, the eruption time is not defined by the dynamics underlying the exponential model, and eruption forecasts based on this model must be based on other metrics, such as earthquake rate or total number thresholds. If forecasts are to be based on either the IOL or exponential models, it is important to be able to distinguish between these competing hypotheses on the basis of the pre-eruptive earthquake data alone.

The Bayesian information criterion (BIC) provides a simple means of choosing the best model based on the likelihood of the observations given the model, with a bias favouring the model with fewer parameters. The BIC has been used widely for pragmatic decision making, as supporting evidence for a model; see Kass & Raftery (1995) for a comprehensive review. The BIC is given by

$$BIC = -2 \ln(L) + P \ln(n), \quad (7)$$

where  $L$  is the likelihood of the observations given the model,  $P$  is the number of free parameters and  $n$  is the number of observations. In making an inference, the preferred model is more likely to have the lower  $BIC$ . Where  $\Delta BIC$  is the positive difference in  $BIC$  between the two models, the difference in likelihood is given by  $\exp(\Delta BIC/2)$  in the case where both models have the same number of parameters. The BIC has previously been shown to be effective at distinguishing between competing IOL and exponential models in retrospective analyses of earthquake rates before basaltic eruptions (Bell *et al.* 2011a; Bell & Kilburn 2012).

Fig. 5 shows  $\Delta BIC$  between the IOL model (1) the ML exponential model (eq. 6) and (2) a simple ML constant rate model ( $\lambda(t) = c$ ) as the pre-eruptive sequence progresses. Negative values indicate a preference for the IOL. Early in the sequence, the  $BIC$  slightly favours both the exponential and constant rate models over the IOL, even though the data are generated from an IOL. This preference is due to the additional parameters in the IOL model (three, rather than two in the exponential model, and one in the constant rate model), and indicates that the difference in likelihood between the models is negligible until close to the eruption. The IOL is, on average, preferred over the constant rate model after just over 300 d. It is not until around 50 d before the eruption that there is a strong

preference for the IOL over an exponential model, the same point at which  $p$  converges, with similar scalability to shorter duration sequences.

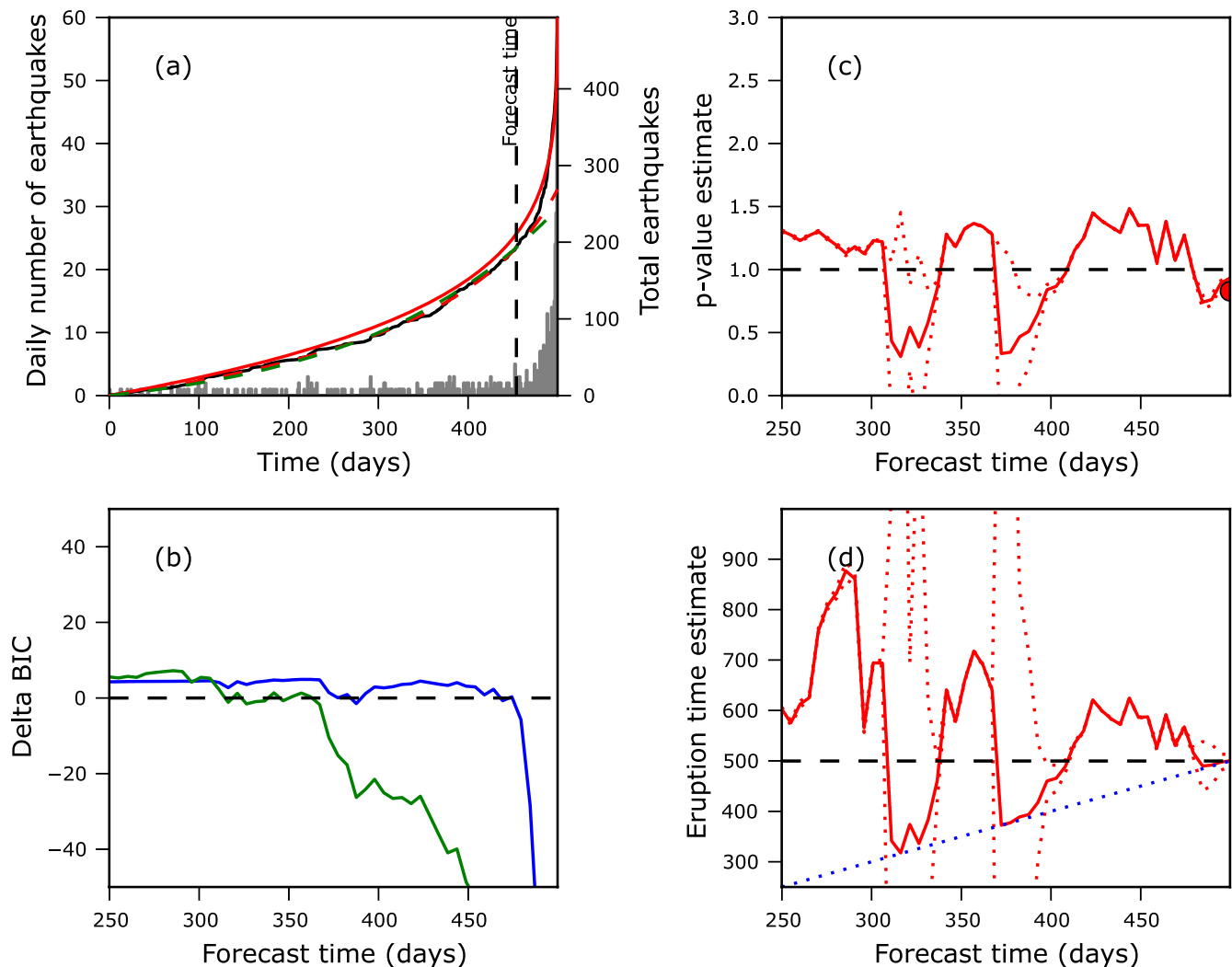
Figs 2–5 describe the behaviour of forecasts based on the IOL when aggregated over many realizations. Fig. 6 illustrates the nature of ML forecasting information for a single realization of the pre-eruptive earthquake data. The raw data, prospective IOL and exponential ML models based on the first 90 per cent of the data, and retrospective IOL model (where the eruption time is known) are shown in Fig. 6(a). The IOL and exponential models are almost indistinguishable at this forecast time. This observation which is supported by the trend in  $\Delta BIC$ ; the IOL is not favoured until after 475 d (Fig. 6b). The ML estimate of both  $p$  (Fig. 6c) and eruption time (Fig. 6d) vary considerably with time and show no simple convergence trend; they are sensitive to the exact occurrence time of earthquake within the sequence. The error in these parameter values, determined using the parameter covariance matrix based on the Fisher information (Ogata 1983), clearly underestimates the true error at many forecast times. It is not until about 10 d before the eruption that the estimated eruption time approaches the true value. Notably, the IOL forecast would result in multiple false alarms (where the eruption time estimate coincides with the time at which the forecast is made, e.g. at 320 and 370 d) and late forecasts (e.g. at 450 d).

#### 4 EXAMPLES OF FORECASTS USING REAL DATA

We now apply the ML forecasting techniques to real earthquake data before a selection of volcanic eruptions where the IOL model has previously been applied. Our aim is to illustrate the uncertainties involved in forecasting on pre-eruptive sequences before different types of eruptions that evolve over different timescales. We also highlight the similarities and differences to model behaviour on idealized data, thus identifying inadequacies of the IOL.

We analyse four different pre-eruptive earthquake catalogues: (1) magnitude 2.5 and greater earthquakes preceding the basaltic





**Figure 6.** Maximum-likelihood eruption forecasting—synthetic data example: (a) daily (grey bars) and total (solid black line) numbers of earthquakes, retrospective ML IOL model (solid red line), and prospective ML IOL and exponential model at 90 per cent of eruption time (dashed red and green lines, respectively); (b) Delta BIC as a function of forecast time for IOL—Exponential (blue) and IOL—Constant rate (green); (c) ML estimate of  $p$ -value for IOL model (solid red line) and uncertainty (dotted red line) as a function of forecast time and retrospective ML  $p$ -value estimate (red semi-circle) and (d) ML estimate of eruption time for IOL model (solid red line) and uncertainty (dotted red line) as a function of forecast time. Dotted blue line has a slope of 1 and is for reference. Note the false alarms at 320 and 370 d.

flank eruption beginning on the 1989 September 24 at Mt Etna (Vinciguerra 2002); (2) magnitude 1.0 and greater earthquakes before a lava dome eruption at Mount St Helens beginning on 1985 May 28 (Malone 1990; Smith *et al.* 2007); (3) magnitude 0.9 and greater earthquakes before a second lava dome eruptions at Mount St Helens beginning on 1986 May 8 (Smith *et al.* 2007) and (4) all earthquakes recorded in the summit cluster before the initial lava dome eruption of Mt Pinatubo on 1991 June 7 (Harlow *et al.* 1996; Hoblitt *et al.* 1996; Smith & Kilburn 2010).

The magnitude thresholds for sequences (1)–(3) are based upon the maximum curvature of the Gutenberg–Richter distribution (Wiemer & Wyss 2000), and potentially represent a slight underestimate of the true completeness magnitude. The frequency magnitude distribution for the Pinatubo sequence does not follow a Gutenberg–Richter distribution, with many earthquakes included that do not have a magnitude defined (Smith & Kilburn 2010). In this instance, we included all recorded earthquakes (reported magnitudes range from 0 to 2). Both these magnitude thresholds and the onset time of the pre-eruptive sequence,  $t_1$ , are determined retrospectively (with

the full data set available). In practice, these criteria will not be known *a priori* and must also be determined from the data as the sequence progresses, potentially increasing the forecast uncertainty.

#### 4.1 Mt Etna—1989 September 24

278 Earthquakes of magnitude 2.5 or greater are reported in the 429 d preceding the 1989 September 24 eruption of Mt Etna, Sicily (Fig. 7a). The BIC analysis indicates a marked preference of the IOL over a constant rate model after 310 d (Fig. 7b). However, there is little difference in BIC between IOL and exponential models for the entire duration of the sequence. The ML estimate of the value of  $p$  is relatively high (around 2.0) except for a period between 300 and 350 d (Fig. 7c). This value is in contrast to the retrospective analysis value (0.6), when the eruption time is known. The forecast eruption time does not converge on the true eruption time, but from 360 d into the sequence it remains 50–100 d late (Fig. 7d). This lateness is increased by a short period of relative quiescence before a final pre-eruptive swarm.

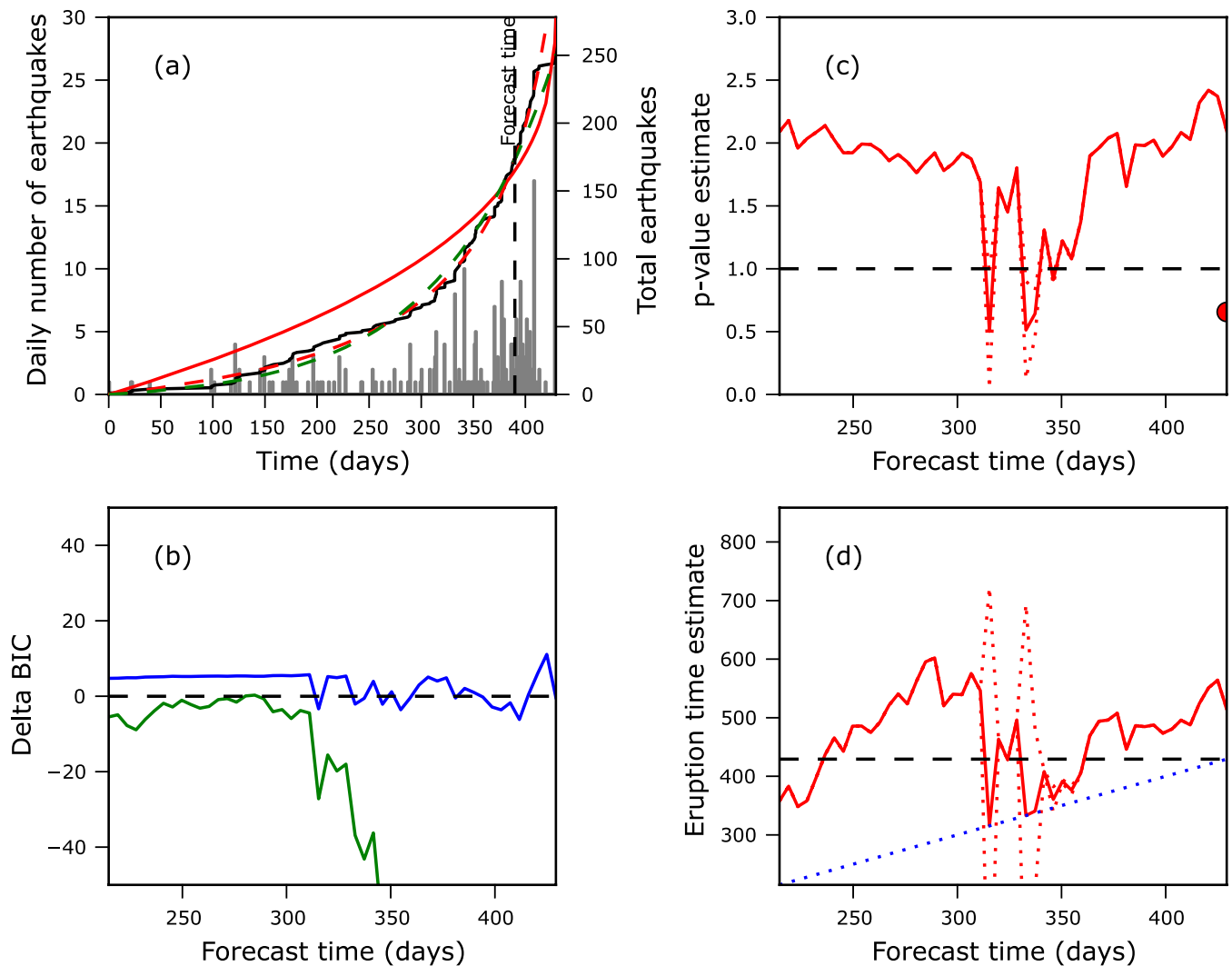


Figure 7. ML eruption forecasting—Mt Etna eruption, 1989 September 24. The composition of panels (a)–(d) and line colour scheme is as for Fig. 6.

#### 4.2 Mount St Helens—1985 May 28

200 earthquakes of magnitude 1.0 or greater are reported in the 214 d preceding the 1985 May 28 eruption of Mount St Helens, USA (Fig. 8a). The BIC analysis indicates a marked preference of the IOL over both a constant rate model and an exponential model after 190 d (Fig. 8b). The ML estimate of the value of  $p$  gradually increases to a value close to 1.0, coinciding with the value obtained in retrospective analysis (Fig. 8c). Likewise, the forecast eruption time converges on the true eruption time (Fig. 8d), although it is clearly influenced by small changes in rate (e.g. at 190 d).

#### 4.3 Mount St Helens—1986 May 8

342 earthquakes of magnitude 1.0 or greater are reported in the 118 d preceding the 1986 May 8 eruption of Mount St Helens (Fig. 9a). The BIC analysis indicates a marked preference of the IOL over both a constant rate model and an exponential model after 100 d (Fig. 9b). There is a temporary increase in the preference for the IOL associated with an increase in the earthquake rate after 75 d. The ML estimate of the value of  $p$  fluctuates around 1.0 between 80 and 110 d. After 110 d, it progressively increases to a value

close to 2.0 (Fig. 9c). This value is in contrast to the retrospective analysis value (0.9), when the eruption time is known. After 100 d, the forecast eruption time converges on the true eruption time, but becomes slightly late in the final few days (Fig. 9d).

#### 4.4 Mt Pinatubo—1991 June 7

340 earthquakes are reported in the 8 d preceding the 1991 June 7 eruption of Mt Pinatubo, Philippines (Fig. 10a). The BIC analysis indicates a marked preference of the IOL over both a constant rate model after 5.5 d (Fig. 10b), and a less marked and variable preference over the exponential model after 6.7 d. The ML estimate of the value of  $p$  fluctuates considerably, and increases rapidly after 7.0 d to a value above 3.0 (Fig. 10c). Again, this value is in contrast to the retrospective analysis value (1.0), determined when the eruption time is known. Until 7.5 d, the forecast eruption time converges on the true eruption time (Fig. 10d). After 7.5 d, it increases to values that are between 2 and 4 d late.

All four real examples share some common similarities. In each case, forecasts based on the IOL would result in several false alarms, and distinguishing between an IOL and exponential model on the basis of the data alone would be difficult until relatively close to

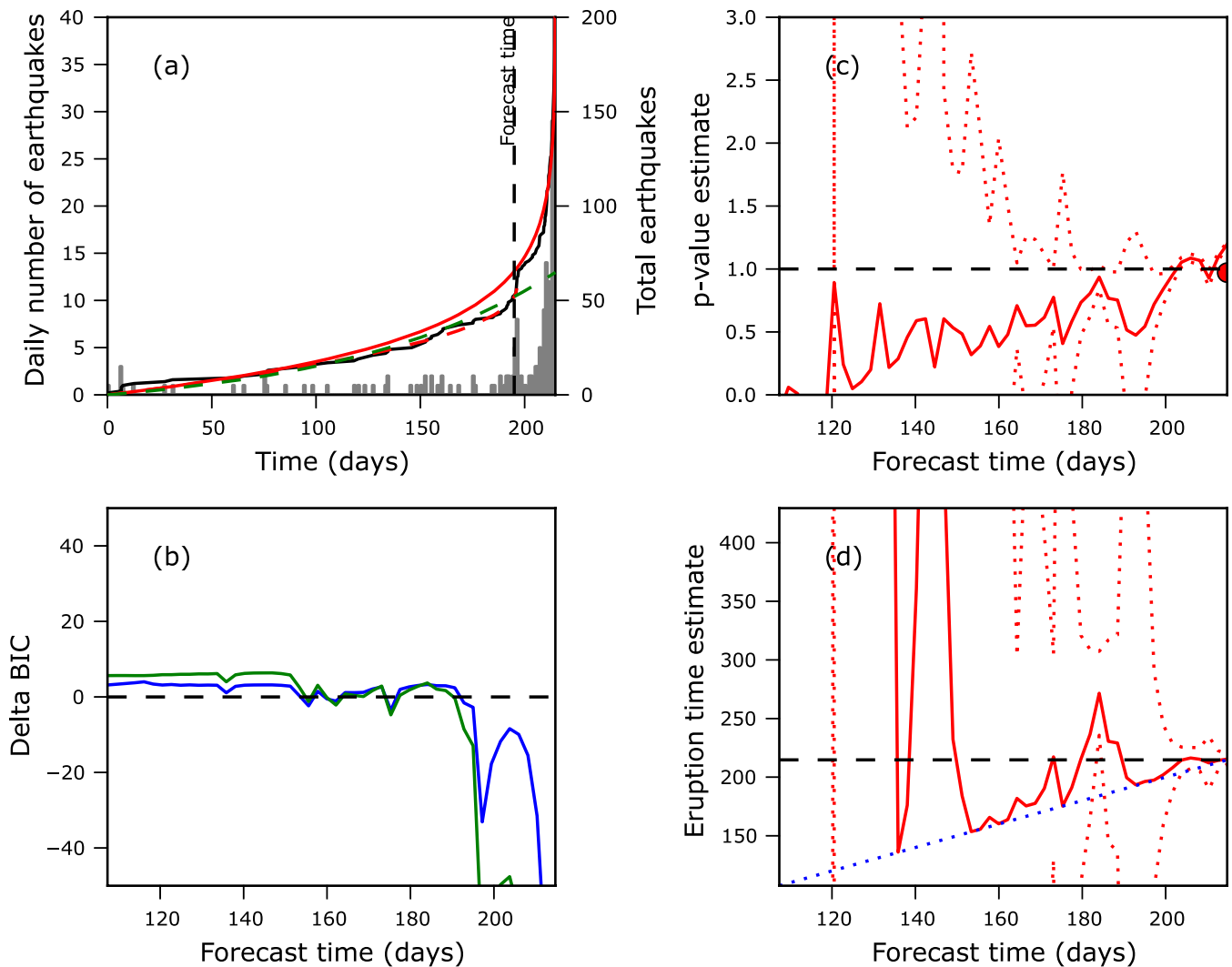


Figure 8. ML eruption forecasting—Mount St Helens eruption, 1985 May 25. The composition of panels (a)–(d) and line colour scheme is as for Fig. 6.

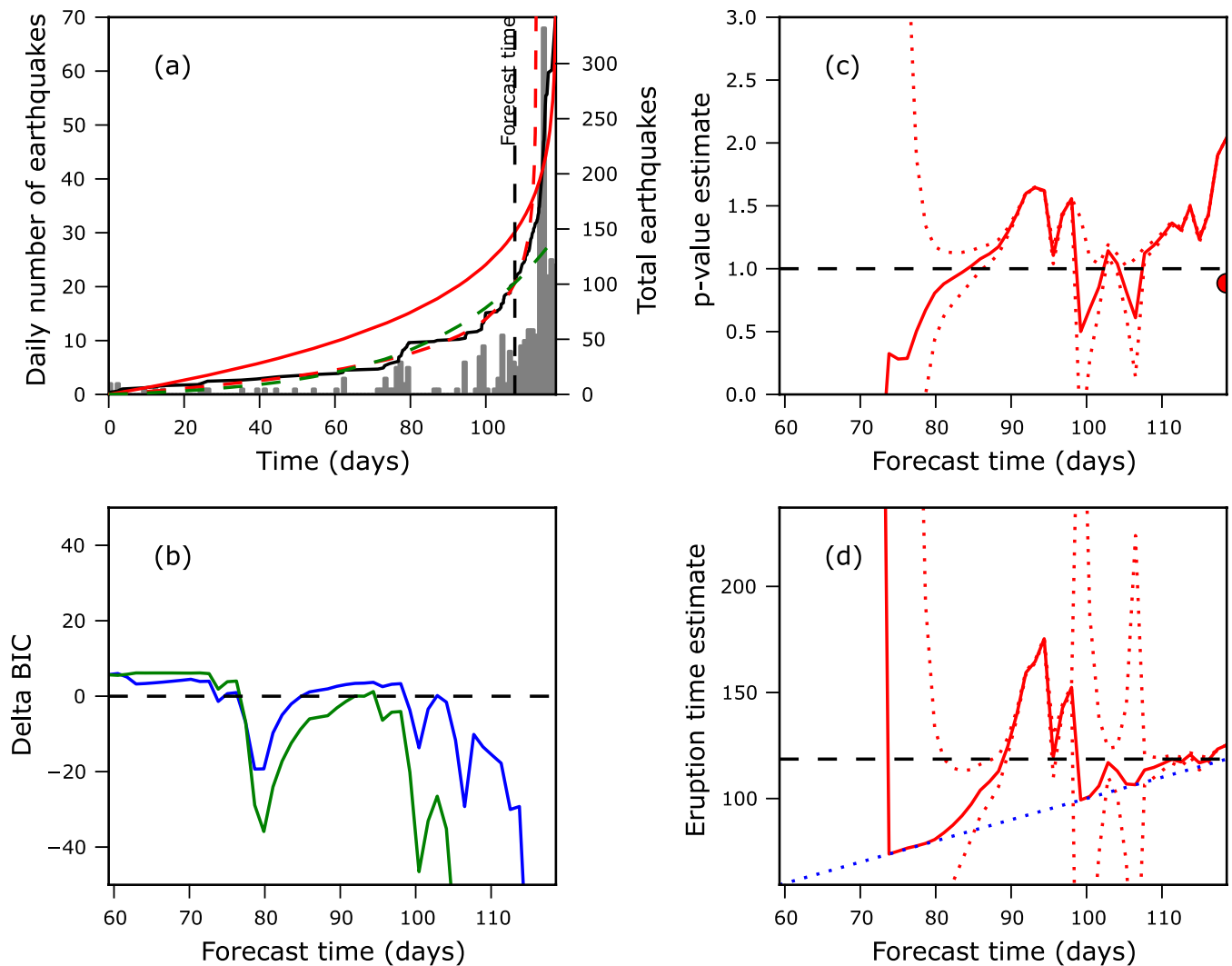
the eruption. Both these properties are entirely consistent with the behaviour observed with the synthetic data, and are inherent problems associated with applying the IOL to earthquake data. All four examples also show a discrepancy between the prospectively and retrospectively estimated value of  $p$ . This behaviour is not expected from the synthetic analyses, and suggests a real deviation from the IOL model very close to the eruption time. The deviation at short times also forces a poor retrospective fit.

## 5 DISCUSSION

The primary aim of this work is to provide statistics to illustrate the type of uncertainties associated with forecasting the onset time of volcanic eruptions based on the IOL. Using simulated data, we are able to consider the ‘best-case’ scenario, where we exclude the possibility of model and data errors, and examine the accuracy and bias of forecast eruption onset times that result from model parameter estimation based on single realizations of stochastic point process time-series. The GLM method has a much increased accuracy compared to the FFM, but forecasts are biased by binning close to the true eruption time. The ML method provides the most accurate and least-biased forecasts, and is reliable when the value

of  $p$  is not known *a priori*. However, even in the condition that we assume we know the true value of the power-law exponent  $p$  *a priori*, we observe a wide range of forecast eruption times for different realizations of the same model, although the mean forecast eruption time for many different realizations are close to the true value. The implication is that, for any single pre-eruptive earthquake sequence, there will inherently be significant uncertainty associated with the estimated time of eruption. The precise values of the uncertainty will depend on the exact parameter values, including the duration of the pre-eruptive sequence and sensitivity of the monitoring network. Higher values of  $p$  result in stronger non-linearity and reduced forecast accuracy; slightly increased forecast accuracy is associated with lower values of  $p$ .

The forecast accuracy decreases considerably when the value of  $p$  also has to be estimated from the data. It is possible that some prior constraint on the value of  $p$  may be established from past data (from the specific volcano of interest, analogue volcanoes or more generically), and theoretical and experimental studies. This analysis would naturally extend into a Bayesian framework where such prior distributions can be taken into account (e.g. Holschneider *et al.* 2012). This finding also highlights the need for better observational and theoretical constraint of parameter values.



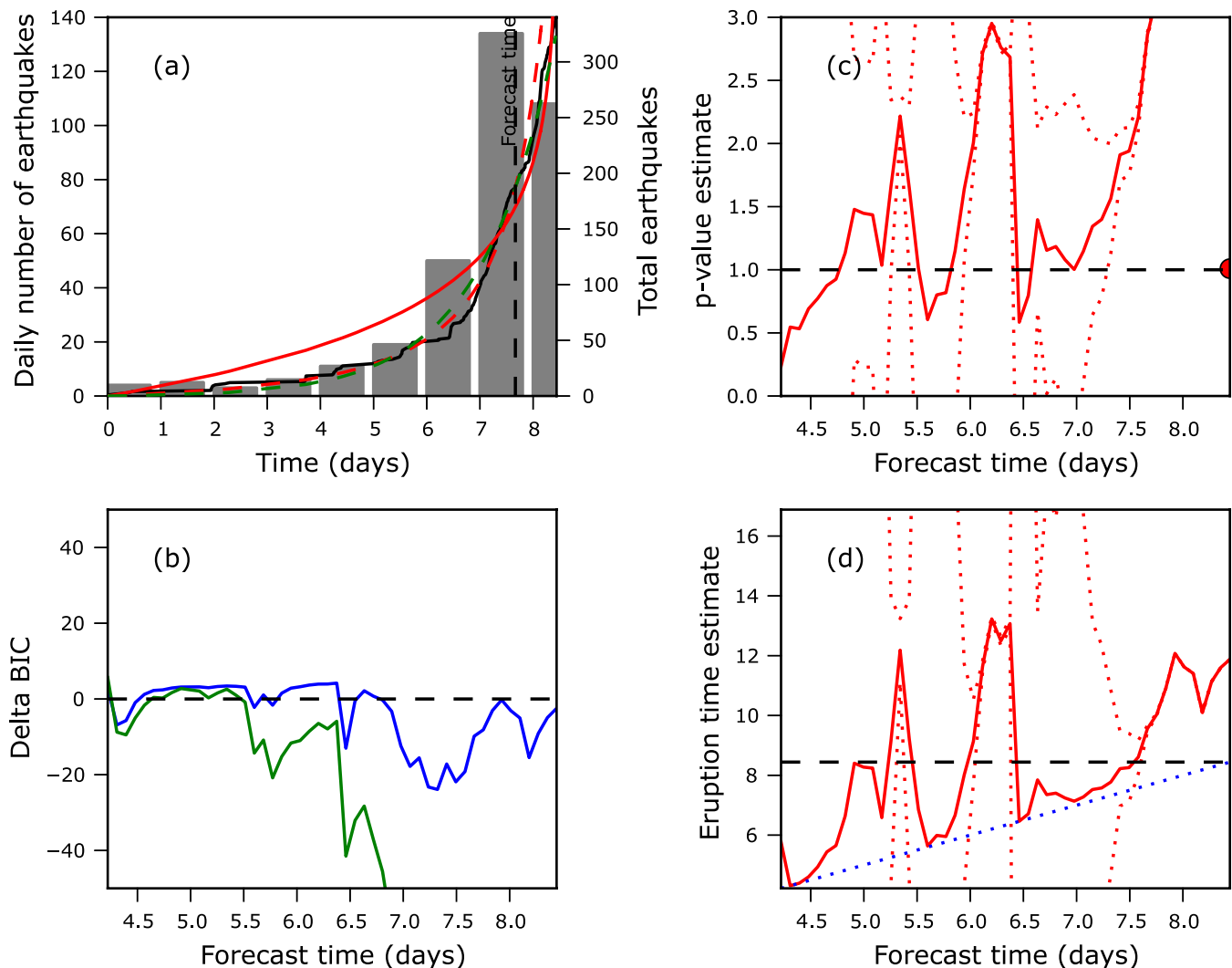
**Figure 9.** ML eruption forecasting—Mount St Helens eruption, 1986 May 8. The composition of panels (a)–(d) and line colour scheme is as for Fig. 6.

Using a limited number of examples it is not possible to evaluate how well the IOL performs on real data. The examples here support the observations made using synthetic data. They show that forecasting based on the IOL would result in both a significant number of false alarms and late forecasts. The prospective ML estimate of  $p$  does not converge to the retrospectively determined value (where the eruption time is known and fixed). The retrospective value is close to 1.0 for the Mount St Helens and Pinatubo examples, and 0.6 for the Mt Etna example. It is also notable that the forecasts from Pinatubo and Mt Etna do not converge to the true eruption time as the eruption is approached. This behaviour is not observed in any of the synthetic examples, suggesting that the real data are deviating from the simple IOL model in these instances. The late forecasts are associated with a drop in earthquake rate just before the eruption, potentially due to either an increase in the completeness magnitude (perhaps due to masking by pre-eruptive tremor), or a deviation from the IOL close to eruption (perhaps indicative of a period of magma ascent). Some form of deviation from the IOL close to the eruption is required to avoid unphysical infinite earthquake rates, and in real scenario may arise due to weak shallow crust, lower confining pressures or the presence of aquifers. Such factors are likely to result in greater forecast uncertainty close to the eruption in real data than in synthetics.

In practice, we expect many factors to contribute to forecast uncertainties that are greater than we demonstrate here. In addition to parameter uncertainty, model and data errors will be an issue in real forecasting scenarios. At best, the IOL will only approximate the true pre-eruptive behaviour at a volcano, and only a fraction of inverse Omori-like sequences will end in an eruption. The influence of additional processes such as background earthquakes, temporally varying magmatic stresses and earthquake triggering are all likely to lead to deviations away from the ideal model behaviour. In addition, real data will be noisy, with magnitude and timing uncertainties, for example. It is possible to include these phenomena in forecasting models. However, additional model complexity and data uncertainty will result in forecasts with a greater uncertainty than those due simply to parameter estimation. Further work, including more sophisticated simulations and an exhaustive study of archive data is needed to begin to quantify these effects.

## 6 CONCLUSIONS

Using synthetic data, we demonstrate the variability of forecasts of the eruption onset time that arises from estimating the parameters of the IOL. Comparing different methodologies, we find that an ML provides the most accurate and least-biased forecasts. When



**Figure 10.** ML eruption forecasting—Mt Pinatubo eruption, 1991 June 7. The composition of panels (a)–(d) and line colour scheme is as for Fig. 6.

analysed over many realizations, the mean forecasts quickly converge to the true solution. However, although the variability in the forecast converges as the eruption time is approached, it remains considerable. For typical parameter values and a 500-d sequence, at 25 d before the eruption 10 per cent of the forecasts are more than 8 d early or late in the case that the power-law exponent is known *a priori*, and more than 18 d early or late if the power-law exponent is unknown. Such values define an upper limit to the predictability of eruption times. False alarms and late forecasts are an inherent feature of the mathematics of the IOL, even on these idealized data sets. Examples from real data suggest that the predictability is much lower in practice, particularly close to the eruption time where the IOL performs less well, indicative of physical processes not captured by the model. Volcanic hazard management strategies must incorporate methods to effectively deal with such uncertainty in forecast eruption times.

#### ACKNOWLEDGEMENTS

We thank Matthias Holschneider and Clement Narteau for helpful discussion, Agust Gudmundsson and Michael West for comments on a previous version of this manuscript and Yan Lavalée and an anonymous reviewer for their constructive reviews. The work was

funded by a UK Natural Environment Research Council standard award EFFORT (Ne/H02297x/1) and a Scottish Government and Royal Society of Edinburgh fellowship (MN).

#### REFERENCES

- Bell, A.F. & Kilburn, C.R.J., 2012. Precursors to dyke-fed eruptions at basaltic volcanoes: insights from patterns of volcano-tectonic seismicity at Kilauea volcano, Hawaii, *Bull. Volc.*, **74**, 325–339.
- Bell, A.F., Greenhough, J., Heap, M.J. & Main, I.G., 2011a. Challenges for forecasting based on accelerating rates of earthquakes at volcanoes and laboratory analogues, *Geophys. J. Int.*, **185**, 718–723.
- Bell, A.F., Naylor, M., Heap, M.J. & Main, I.G., 2011b. Forecasting volcanic eruptions and other material failure phenomena: an evaluation of the failure forecast method, *Geophys. Res. Lett.*, **38**, doi:10.1029/2011gl048155.
- Chouet, B.A. & Matoza, R.S., 2013. A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption, *J. Volc. Geotherm. Res.*, **252**, 108–175.
- Cornelius, R.R. & Voight, B., 1994. Seismological aspects of the 1989–1990 eruption at Redoubt Volcano, Alaska—the materials failure forecast method (FFM) with RSAM and SSAM seismic data, *J. Volc. Geotherm. Res.*, **62**, 469–498.
- Daley, D.J. & Vere-Jones, D., 2005. *An Introduction to the Theory of Point Processes*, 2nd edn, Vol. 1, Springer, New York.



- De la Cruz-Reyna, S. & Reyes-Davila, G.A., 2001. A model to describe precursory material-failure phenomena: applications to short-term forecasting at Colima volcano, Mexico, *Bull. Volc.*, **63**, 297–308.
- Gudmundsson, A., 2006. How local stresses control magma-chamber ruptures, dyke injections, and eruptions in composite volcanoes, *Earth-Sci. Rev.*, **79**, 1–31.
- Gudmundsson, A., 2012. Magma chambers: formation, local stresses, excess pressures, and compartments, *J. Volc. Geotherm. Res.*, **237**, 19–41.
- Hammer, C. & Neuberg, J.W., 2009. On the dynamical behaviour of low-frequency earthquake swarms prior to a dome collapse of Soufriere Hill volcano, Montserrat, *Geophys. Res. Lett.*, **36**, doi:10.1029/2008gl036837.
- Harlow, D.H., Power, J.A., Laguerta, E.P., Ambubuyog, G. & Hoblitt, R.P., 1996. Precursory seismicity and forecasting of the June 15, 1991, eruption of Mount Pinatubo, in *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 223–247, eds Punongbayan, R.S. & Newhall, C.G., University of Washington Press, Seattle.
- Hoblitt, R.P., Mori, J. & Power, J.A., 1996. Computer visualization of earthquake hypocentres, in *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, pp. 383–385, eds Punongbayan, R.S. & Newhall, C.G., University of Washington Press, Seattle.
- Holschneider, M., Narteau, C., Shebalin, P., Peng, Z. & Schorlemmer, D., 2012. Bayesian analysis of the modified Omori law, *J. geophys. Res.-Solid Earth*, **117**, doi:10.1029/2011jb009054.
- Kass, R.E. & Raftery, A.E., 1995. Bayes factors, *J. Am. Stat. Assoc.*, **90**, 773–795.
- Kilburn, C., 2012. Precursory deformation and fracture before brittle rock failure and potential application to volcanic unrest, *J. geophys. Res.-Solid Earth*, **117**, doi:10.1029/2011jb008703.
- Kilburn, C.R.J., 2003. Multiscale fracturing as a key to forecasting volcanic eruptions, *J. Volc. Geotherm. Res.*, **125**, 271–289.
- Kilburn, C.R.J. & Voight, B., 1998. Slow rock fracture as eruption precursor at Soufriere Hills volcano, Montserrat, *Geophys. Res. Lett.*, **25**, 3665–3668.
- Lavallee, Y., Meredith, P.G., Dingwell, D.B., Hess, K.U., Wassermann, J., Cordonnier, B., Gerik, A. & Kruhl, J.H., 2008. Seismogenic lavas and explosive eruption forecasting, *Nature*, **453**, 507–510.
- Lavallee, Y., Benson, P.M., Heap, M.J., Hess, K.U., Flaws, A., Schillinger, B., Meredith, P.G. & Dingwell, D.B., 2013. Reconstructing magma failure and the degassing network of dome-building eruptions, *Geology*, **41**, 515–518.
- Lengliné, O., Marsan, D., Got, J.L., Pinel, V., Ferrazzini, V. & Okubo, P.G., 2008. Seismicity and deformation induced by magma accumulation at three basaltic volcanoes, *J. geophys. Res.-Solid Earth*, **113**, B12305, doi:10.1029/2008JB005937.
- Main, I.G., 1999. Applicability of time-to-failure analysis to accelerated strain before earthquakes and volcanic eruptions, *Geophys. J. Int.*, **139**, F1–F6.
- Main, I.G., 2000. A damage mechanics model for power-law creep and earthquake aftershock and foreshock sequences, *Geophys. J. Int.*, **142**, 151–161.
- Malone, S.D., 1990. Mount St Helens, the 1980 re-awakening and continuing seismic activity, *Geosci. Can.*, **17**, 146–150.
- Marzocchi, W. & Bebbington, M.S., 2012. Probabilistic eruption forecasting at short and long time scales, *Bull. Volc.*, **74**, 1777–1805.
- McNutt, S.R., 2002. Volcano seismology and monitoring for eruptions, in *International Handbook of Earthquake & Engineering Seismology*, pp. 383–406, eds Lee, W.H.K., Kanamori, H., Jennings, P.C. & Kisslinger, C., IASPEI, Palo Alto, CA.
- McNutt, S.R., 2005. Volcanic seismology, *Annu. Rev. Earth planet. Sci.*, **33**, 461–491.
- Murphy, A.H., 1993. What is a good forecast—an essay on the nature of goodness in weather forecasting, *Weather Forecast.*, **8**, 281–293.
- Nelder, J.A. & Wedderburn, R.W.M., 1972. Generalized Linear Models, *J. R. Stat. Soc. A*, **135**, 370–384.
- Neuberg, J.W., Tuffen, H., Collier, L., Green, D., Powell, T. & Dingwell, D., 2006. The trigger mechanism of low-frequency earthquakes on Montserrat, *J. Volc. Geotherm. Res.*, **153**, 37–50.
- Ogata, Y., 1983. Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum-likelihood procedure, *J. Phys. Earth*, **31**, 115–124.
- Reasenber, P.A. & Jones, L.M., 1989. Earthquake hazard after a mainshock in California, *Science*, **243**, 1173–1176.
- Reyes-Davila, G.A. & De la Cruz-Reyna, S., 2002. Experience in the short-term eruption forecasting at Volcan de Colima, Mexico, and public response to forecasts, *J. Volc. Geotherm. Res.*, **117**, 121–127.
- Roman, D.C. & Cashman, K.V., 2006. The origin of volcano-tectonic earthquake swarms, *Geology*, **34**, 457–460.
- Schmid, A. & Grasso, J.R., 2012. Omori law for eruption foreshocks and aftershocks, *J. geophys. Res.-Solid Earth*, **117**, doi:10.1029/2011jb008975.
- Smith, R. & Kilburn, C.R.J., 2010. Forecasting eruptions after long repose intervals from accelerating rates of rock fracture: the June 1991 eruption of Mount Pinatubo, Philippines, *J. Volc. Geotherm. Res.*, **191**, 129–136.
- Smith, R., Kilburn, C.R.J. & Sammonds, P.R., 2007. Rock fracture as a precursor to lava dome eruptions at Mount St Helens from June 1980 to October 1986, *Bull. Volc.*, **69**, 681–693.
- Sparks, R.S.J., 2003. Forecasting volcanic eruptions, *Earth planet. Sci. Lett.*, **210**, 1–15.
- Tokarev, P.I., 1971. Forecasting volcanic eruptions from seismic data, *Bull. Volc.*, **35**, 243–250.
- Tuffen, H., Smith, R. & Sammonds, P.R., 2008. Evidence for seismogenic fracture of silicic magma, *Nature*, **453**, 511–514.
- Vinciguerra, S., 2002. Damage mechanics preceding the September–October 1989 flank eruption at Mount Etna volcano inferred by seismic scaling exponents, *J. Volc. Geotherm. Res.*, **113**, 391–397.
- Voight, B., 1988. A method for prediction of volcanic-eruptions, *Nature*, **332**, 125–130.
- Wiemer, S. & Wyss, M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan, *Bull. seism. Soc. Am.*, **90**, 859–869.